

Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWTID1720001058
Project Title	Panic Disorder Detection
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	<p>NaN values were found out in the data set</p> <pre>data.isnull().sum() Participant ID 0 Age 0 Gender 0 Family History 0 Personal History 0 Current Stressors 0 Symptoms 0 Severity 0 Impact on Life 0 Demographics 0 Medical History 25173 Psychiatric History 24921 Substance Use 33374 Coping Mechanisms 0 Social Support 0 Lifestyle Factors 0 Panic Disorder Diagnosis 0 dtype: int64</pre>	Low	<p>Removed through dropping the columns</p> <pre>data.drop(columns=['Medical History', 'Psychiatric History'], inplace = True)</pre>

<p>Dataset</p>	<p>More categorical data</p>	<p>Converted to numeric type by label encoding technique</p>
	<p>high</p>	<pre> from sklearn.preprocessing import LabelEncoder label_encoder = LabelEncoder() # Encoding the categorical data and converting them into numeric by label encoding technique columns_to_encode = ["Gender", "Family History", "Personal History", "Current Stressors", "Symptoms", "Severity", "Impact on Life", "Demographics", "Substance Use", "Coping Mechanisms", "Social Support", "Lifestyle Factors", "Patient Education"] # encode each value for col in columns_to_encode: data[col] = label_encoder.fit_transform(data[col].astype(str)) # display the transformed dataset data.head() </pre>